

Genomic selection across multiple breeding cycles in applied bread wheat breeding

Sebastian Michel¹ · Christian Ametz² · Huseyin Gungor³ · Doru Epure⁴ · Heinrich Grausgruber⁵ · Franziska Löschnerberger² · Hermann Buerstmayr¹

Received: 5 October 2015 / Accepted: 15 January 2016 / Published online: 11 April 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract

Key message We evaluated genomic selection across five breeding cycles of bread wheat breeding. Bias of within-cycle cross-validation and methods for improving the prediction accuracy were assessed.

Abstract The prospect of genomic selection has been frequently shown by cross-validation studies using the same genetic material across multiple environments, but studies investigating genomic selection across multiple breeding cycles in applied bread wheat breeding are lacking. We estimated the prediction accuracy of grain yield, protein content and protein yield of 659 inbred lines across

five independent breeding cycles and assessed the bias of within-cycle cross-validation. We investigated the influence of outliers on the prediction accuracy and predicted protein yield by its components traits. A high average heritability was estimated for protein content, followed by grain yield and protein yield. The bias of the prediction accuracy using populations from individual cycles using fivefold cross-validation was accordingly substantial for protein yield (17–712 %) and less pronounced for protein content (8–86 %). Cross-validation using the cycles as folds aimed to avoid this bias and reached a maximum prediction accuracy of $r_{GS} = 0.51$ for protein content, $r_{GS} = 0.38$ for grain yield and $r_{GS} = 0.16$ for protein yield. Dropping outlier cycles increased the prediction accuracy of grain yield to $r_{GS} = 0.41$ as estimated by cross-validation, while dropping outlier environments did not have a significant effect on the prediction accuracy. Independent validation suggests, on the other hand, that careful consideration is necessary before an outlier correction is undertaken, which removes lines from the training population. Predicting protein yield by multiplying genomic estimated breeding values of grain yield and protein content raised the prediction accuracy to $r_{GS} = 0.19$ for this derived trait.

Communicated by J. Reif.

S. Michel and C. Ametz contributed equally to the work.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-016-2694-2) contains supplementary material, which is available to authorized users.

✉ Hermann Buerstmayr
hermann.buerstmayr@boku.ac.at

- ¹ Department for Agrobiotechnology (IFA-Tulln), Institute for Biotechnology in Plant Production, University of Natural Resources and Life Sciences, Vienna (BOKU), Konrad-Lorenz-Str. 20, 3430 Tulln, Austria
- ² Saatzucht Donau GesmbH. and CoKG, Saatzuchtstrasse 11, 2301 Probstdorf, Austria
- ³ ProGen Seed A.Ş., Büyükdalyan Mah. 2. Kime evler Sok., No: 49, Antakya/Hatay, Turkey
- ⁴ Probstdorfer Saatzucht Romania SRL, Str. Sirlui Nr.20, sect. 1, Bucharest, Romania
- ⁵ Plant Breeding Division, Department of Crop Science, University of Natural Resources and Life Sciences, Vienna (BOKU), Konrad-Lorenz-Str. 24, 3430 Tulln, Austria

Introduction

Originally suggested by Meuwissen et al. (2001), genomic selection showed great promise to strongly increase the rate of genetic improvement in both animal and plant breeding programs. This new method allowed a comparative larger gain from selection by estimating all marker effects simultaneously and subsequent selection of genetically superior individuals based on their genomic estimated breeding value (GEBV) (Bernardo and Yu 2007; Piyasatian

et al. 2007), instead of using a few significant markers as in classical marker-assisted selection (Lande and Thompson 1990). Genomic selection was readily integrated into applied animal breeding programs due to a high accuracy of breeding values and a previously existing similar system, which employed pedigree instead of marker information (VanRaden 2008; Hayes et al. 2009). Notwithstanding, the relative simple methodology made genomic selection also interesting for applied plant breeding: a training population of phenotyped and genotyped individuals is utilized to establish a statistical model that predicts breeding values of non-phenotyped individuals from a selection or validation population by their genomic fingerprints (Heffner et al. 2009; Jannink et al. 2010).

Although methodologically simple the sparse knowledge about its functionality made it initially difficult to find starting points for increasing the prediction accuracy. Theoretical studies thus laid the foundation for optimizing breeding with genomic selection by trying to understand the underlying mechanics of this ‘green box’ approach. The driving forces of prediction accuracy that can be most readily influenced by plant breeders are the training population size and heritability (Muir 2007; Hayes et al. 2009), by adequately adjusting the resource allocation (Riedelsheimer and Melchinger 2013; Longin et al. 2015). Recent advances in sequencing technologies made it possible to apply cost effective genotyping methods such as genotyping-by-sequencing (GBS) in various crop species (Elshire et al. 2011; Poland et al. 2012; Huang et al. 2014) yielding an appropriate large number of markers for genomic selection (Hayes et al. 2009; Schulz-Streeck et al. 2011). The use of dense genome-wide markers increases the chance of markers being in linkage disequilibrium (LD) with QTL influencing the trait of interest (e.g. Meuwissen et al. 2001), and determines to some extent how well genetic relationship and genetic architecture are captured by the genomic selection model (Daetwyler et al. 2010; Heslot et al. 2013a). The importance of a close genetic relationship between training and validation populations to achieve a high prediction accuracy (Habier et al. 2013) has been verified numerous times in plant breeding studies, e.g. with sugar beet (Würschum et al. 2013); rapeseed (Würschum et al. 2014), maize (Zhao et al. 2012; Riedelsheimer et al. 2013; Albrecht et al. 2014; Lehermeier et al. 2014), and wheat (Charmet et al. 2014; Crossa et al. 2014), which motivated investigations for an optimal training population construction to reduce phenotyping costs (Rincent et al. 2012; Isidro et al. 2015).

Summing up, valuable insights into genomic selection have been gained in relative short time opening up numerous possibilities for its implementation into the framework of plant breeding (Heslot et al. 2015). Notwithstanding, most studies were conducted with the same genetic

material across multiple environments or made use of historical phenotypic data (Dawson et al. 2013; Storlie and Charmet 2013; Ly et al. 2013; Rutkoski et al. 2015), while few have focused on the problem of predicting across breeding cycles in applied plant breeding programs. This issue was addressed for the first time in sugar beet breeding, where genomic selection showed great promise across two subsequent breeding cycles especially for highly heritable traits (Hofheinz et al. 2012). A population of parental lines was employed to predict GEBVs for five successive years in a recent study with six-row barley by Sallam et al. (2015), who reported varying degrees of prediction accuracy depending both on the progeny set and trait. We are not aware of any studies investigating genomic selection across multiple breeding cycles in applied bread wheat breeding; thus the objectives of this study were (1) to estimate the accuracy when predicting grain yield, protein content and protein yield of wheat inbred lines across multiple independent breeding cycles; (2) compare within-cycle and between-cycle prediction accuracy obtained from different cross-validation schemes; and (3) investigate model independent possibilities to increase the prediction accuracy.

Materials and methods

Plant material and phenotypic data

We analyzed five breeding cycles from a commercial winter wheat (*Triticum aestivum* L.) breeding program, where breeding cycles correspond to the different starting years. A total of 659 genotyped lines from multiple families, either $F_{4,6}$ or directly derived by the double haploid method, were tested in multi-environment trials from 2010 to 2014. A balanced subset of trial locations was selected for evaluating the merit of genomic selection across multiple breeding cycles. Within each breeding cycles a different set of 64–176 lines was tested orthogonally across all trial locations. Grain yield (dt ha^{-1}), protein content (%) and protein yield (dt ha^{-1}) were assessed in 2–8 trials per breeding cycles. Protein content was determined by near infrared spectroscopy (NIRS) directly at harvest and protein yield was derived by multiplication of grain yield and protein content on a plot basis. An additional independent set of 178 genotyped lines from the same breeding population was tested in 2015 employing the same phenotyping procedure as in 2010–2014. Trial locations spanned from Austria over Serbia, Croatia, Hungary, and Romania to the Central Anatolian High Plateau in Turkey, thus covering a large diversity of target environments. Trials were designed according to the standard procedure in plant breeding, where non-replicated earlier generation lines are tested

along with replicated checks, which allowed correcting for spatial field trends and estimation of error variances.

Phenotypic analysis

We followed a two-stage analysis strategy of the phenotypic data, where each individual trial, i.e. location by breeding cycle combination was analyzed separately in the first stage. A baseline model without correction for spatial trend was compared by Akaike's Information Criterion (AIC) with models correcting for row and/or column effects, if feasible an autoregressive variance–covariance structure of the residuals was additionally integrated (Burgueno et al. 2000). The model with the smallest AIC was chosen to derive best linear unbiased estimates (BLUE) for each trial. The heritability was computed as suggested by Piepho and Möhring (2007) by $h^2 = \sigma_G^2 / (\sigma_G^2 + \frac{1}{2}MVD)$, where σ_G^2 designates the genetic variance and MVD the mean variance of a difference of the BLUEs. The analysis at the first stage contained both checks and genotyped lines.

We retained only trials with a heritability larger than 0.3 and genotyped lines for the analysis across trials at the second stage. A linear mixed model of the form

$$y_{ij} = \mu + g_i + t_j + gt_{ij} + e \quad (1)$$

was fitted for all traits, where y_{ij} are the BLUEs from the first stage, μ is the grand mean, and g_i is the effect of the i th inbred line. The effect of the j th trial t_j was fixed, while the line by trial interaction effect gt_{ij} was random. The inverse of the squared standard errors of the means derived from the first stage of analysis were used as weights in this stage to take the varying accuracy of phenotypic records into account (Möhring and Piepho 2009). The residual variance was fixed to 1 for grain yield and 0.1 for protein content and protein yield, which allowed the separation of residual and line by trial interaction variances. Heritability estimates across trials were computed using the standard formula $h^2 = \sigma_G^2 / (\sigma_G^2 + t^{-1}\sigma_{GT}^2)$, dividing the line by trial interaction variance σ_{GT}^2 by the number of trials t . All phenotypic analyses were conducted using the statistical package ASReml 3 (VSN International 2015) for the R programming environment (R development core team 2015).

Genotypic data

Leaves for DNA extraction were sampled from $F_{4.5}$ plants in small observation plots after phenotypic pre-selection during early summer. DNA was extracted following the protocol by Saghai-Maroo et al. (1984). All 659 lines were genotyped with approximately 20 K SNP markers using the DarT genotyping-by-sequencing (GBS) approach (Diversity Array Technologies, 2015). Quality control was applied by filtering out markers with a call rate lower than

90 %, a minor allele frequency smaller than 0.05, and more than 10 % of missing data. Missing data of the remaining 9.5K SNP markers was imputed by the MVN-EM algorithm by Poland et al. (2012) which was specially designed for the use of GBS markers.

Genomic selection and genetic relationship

Genomic predictions of inbred lines were estimated using a ridge regression best linear unbiased prediction (RR-BLUP) model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (2)$$

where \mathbf{y} is an $N \times 1$ vector of BLUEs obtained in the phenotypic analysis, \mathbf{b} is a vector of F fixed effects and \mathbf{X} its corresponding $N \times F$ design matrix. \mathbf{Z} is a $N \times M$ matrix, which coded the M markers as either +1 or −1 for homozygous loci and 0 for heterozygous loci. Random marker effects were assumed to follow a normal distribution $\mathbf{u} \sim N(0, \mathbf{I}\sigma_u^2)$ and equally shrunken towards zero given the penalty parameter $\lambda^2 = \sigma_e^2 / \sigma_u^2$ where σ_e^2 is the variance of the residuals which followed $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$. The kinship between lines was estimated by the genomic relationship matrix, which was computed according to Endelman and Jannick (2012):

$$\mathbf{K} = \mathbf{W}\mathbf{W}^T / 2\sum (p_k - 1)p_k \quad (3)$$

where \mathbf{W} is a centered $N \times M$ marker matrix of the i lines with $W_{ik} = Z_{ik} - 2p_k$ and p_k being the allele frequency at the k th locus. The derived variance–covariance matrix was used to fit a genomic best linear unbiased prediction (G-BLUP) model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{g} + \mathbf{e} \quad (4)$$

where \mathbf{g} is an $N \times 1$ vector of genotypic effects with $\mathbf{g} \sim N(0, \mathbf{K}\sigma_G^2)$. Model (4) has been shown to be equivalent to (2) (VanRaden 2008; Piepho 2009) and allowed estimating the accuracy of each individual line by $r_{PEV} = \sqrt{1 - (PEV / \mathbf{G}_{ii}\sigma_G^2)}$ where PEV is the prediction error variance, σ_G^2 the genetic variance explained by the model and \mathbf{G}_{ii} is the diagonal element of the genomic relationship matrix for each line i (Clark et al. 2012). All models for genomic selection were implemented with the R package rrBLUP (Endelman 2011).

Validation and accuracy

At first we estimated the correlation between the accuracy of each individual line r_{PEV} and the genetic relationship to investigate this important driving force of prediction accuracy across several cycles of wheat breeding. The average genetic relationship of the most related lines from the training population was computed for each line in the validation

population and correlated with r_{PEV} . The number of most related lines was varied between 1 and 500, and one breeding cycle was left out at a time using all other breeding cycles as training population. A fixed year effect was included into model (4) to account for the different yield levels of the studied breeding cycles. Prediction accuracy is generally defined as the Pearson correlation between predicted and true breeding values $r_{MT} = r_{GEBV, TBV}$. The true breeding values were unknown in our study; so we estimated prediction accuracy as the correlation between predicted and observed line performance $r_{GS} = r_{GEBV, BLUE}$. Across-cycle prediction accuracy was subsequently assessed by computing marker effect estimates with the RR-BLUP model of all possible pair-wise training and validation population combinations of the five breeding cycles. Three cross-validation schemes each with fivefolds and 100 replicates were employed to cover different aspects of the prediction accuracy r_{GS} :

1. Within-cycle prediction accuracy was computed by randomly dividing the data into equally sized folds using 80 % of lines within each breeding cycle as a training population and subsequent prediction of the left-out fold. This procedure was repeated for every fold and the resulting prediction accuracy was averaged for each of the 100 replicates.
2. The same training populations as in the within-cycle cross-validation were used to separately predict lines of each other breeding cycle. The average prediction accuracy was saved and utilized to estimate the bias of within-cycle versus between-cycle cross-validation.
3. Fivefold cross-validation, where the breeding cycles constituted the folds, was used to estimate the prediction accuracy across cycles. An equal number of lines were randomly sampled from each breeding cycle, simulating a breeding scenario where training populations for genomic selection models are an assembly of several mixed populations from multiple breeding cycles. Training population sizes varied between 16 and 256 lines. An additional fixed year effect was added to model (2) in order to account for the different yield and protein levels in 2010 to 2014.

Furthermore we studied two possibilities for increasing the prediction accuracy across breeding cycles. First outliers were identified by approximating the genetic correlation among environments by their pair-wise prediction accuracies (Heslot et al. 2013b), and breeding cycles or trials with a strongly deviating character were dropped from the training population. The influence of these outliers was subsequently investigated by comparing the prediction accuracy with the full and outlier corrected dataset, using the same across-cycle cross-validation approach as before.

Training population size was kept constant by randomly choosing additional lines from each remaining breeding cycle in the outlier corrected cross-validation. In addition to cross-validation, the 178 lines from 2015 served as an independent validation population and were predicted by estimated marker effects using either the full or outlier corrected dataset.

Finally we investigated the possibility to increase the prediction accuracy of the derived trait protein yield by multiplying GEBVs of its component traits grain yield and protein yield. The prediction accuracy was estimated by the above described across-cycle cross-validation approach with 100 replicates for each training population size.

Results

Quantitative-genetic parameters

The plant material was tested in a broad spectrum of environments ranging from the Pannonian Basin to the Central Anatolian High Plateau. Despite the expected large genotype by environments interaction we observed a medium to high heritability in each individual breeding cycle for grain yield and protein content (Table 1). A relatively large number of trials having at least a heritability larger than 0.3 were pre-selected for this study to achieve valid and robust results. The excellent data quality was also reflected by the medium to high heritability for protein yield in all but one breeding cycle. Estimates of heritability were lower for protein yield than grain yield except for 2013, where it was 26 % larger. The protein content had on average the highest heritability followed by grain yield and protein yield.

Genetic relationship and prediction accuracy of genomic selection

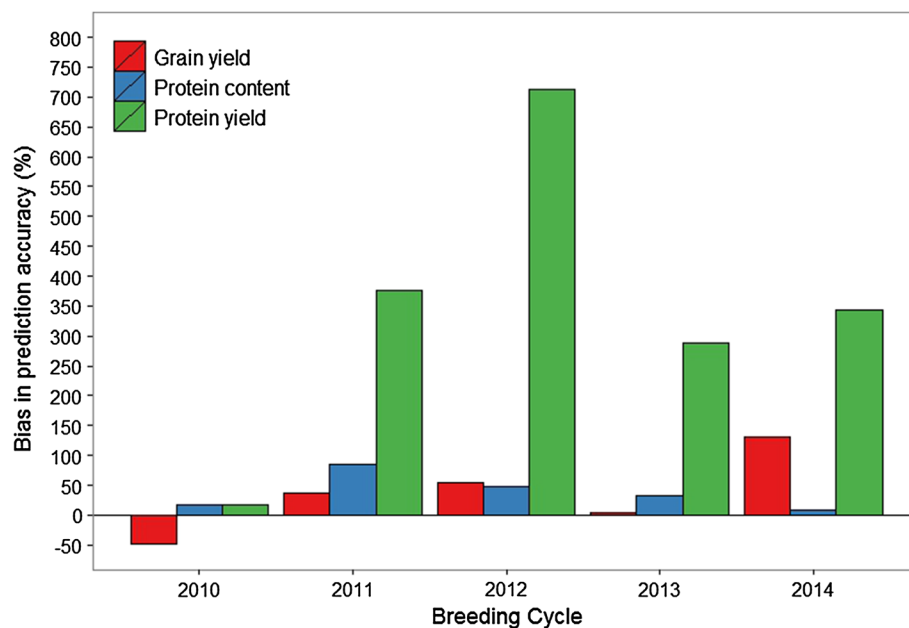
The correlation between accuracy of each individual line and the genetic relationship was strongly dependent on the number of most related lines and the respective validation population (Fig. S1). Optimal correlations for grain yield were achieved using the 11–133 most related lines, while choosing the 70 most related lines led to significant correlations larger than $r = 0.80$ for all validation populations. Similar patterns were observed for protein content and protein yield. The average of the top 70 genetic relationship between lines range from 0.08 to 0.14 within the years, and was smaller between years with an overall average genetic correlation of 0.07 (Fig. S2).

Within-cycle prediction accuracy was compared to between-cycle prediction accuracy by fivefold cross-validation utilizing the same training populations for each

Table 1 Mean, variance components and heritability for grain yield (dt ha⁻¹), protein content (%) and protein yield (dt ha⁻¹) of genotyped lines across all trials in the respective breeding cycles 2010–2014

Trait	Parameter	Breeding cycles				
		2010	2011	2012	2013	2014
Grain yield	Trials	5	6	4	5	8
	σ_G^2	2.28 ± 1.28	4.60 ± 1.60	5.03 ± 1.25	6.64 ± 1.76	37.00 ± 4.71
	σ_{GT}^2	23.70 ± 1.83	23.67 ± 1.99	17.80 ± 1.21	40.98 ± 2.36	54.48 ± 2.26
	h^2	0.32	0.54	0.53	0.45	0.84
Protein content	Trials	4	2	3	4	2
	σ_G^2	0.23 ± 0.05	0.18 ± 0.05	0.35 ± 0.05	0.37 ± 0.06	0.33 ± 0.09
	σ_{GT}^2	0.36 ± 0.04	0.07 ± 0.03	0.27 ± 0.03	0.65 ± 0.05	0.65 ± 0.08
	h^2	0.72	0.84	0.80	0.69	0.50
Protein yield	Trials	4	2	4	4	3
	σ_G^2	0.04 ± 0.03	0.03 ± 0.07	0.05 ± 0.02	0.26 ± 0.05	0.76 ± 0.14
	σ_{GT}^2	0.41 ± 0.04	0.38 ± 0.09	0.34 ± 0.03	0.69 ± 0.05	1.30 ± 0.11
	h^2	0.30	0.14	0.37	0.60	0.64
	Lines	94	64	165	160	176

Genotypic variance (σ_G^2), genotype by trial interaction variance (σ_{GT}^2), and heritability (h^2)

Fig. 1 Bias of the within-cycle prediction accuracy in comparison with the between-cycle prediction accuracy for grain yield, protein content and protein yield and using lines from the years 2010–2014 as training populations

cross-validation scheme. A strong upward bias of within-cycle prediction accuracy was observed for 10 out of 15 traits by cycle combinations and was less than 25 % in four instances (Fig. 1). The bias was especially pronounced for 2014, where the predictive ability of grain yield was overestimated by 130 % and even more for protein yield by 344 %. Protein yield had overall the largest bias ranging from 17 % up to 712 %, while the prediction accuracy of protein content was maximally overestimated by 86 %. Within-cycle cross-validation underestimated the prediction accuracy for grain yield by 47 % merely in one case.

The intention behind using the breeding cycles as folds in a fivefold cross-validation was the avoidance of this bias when estimating the prediction accuracy. Sampling an equal number of lines from each breeding cycle furthermore aimed to avoid a confounding effect between training population sizes and breeding cycle. According to expectation the prediction accuracy increased with the number of lines in the training population (Fig. S3). A maximum was reached for a training population size of 240 lines at $r_{GS} = 0.51$ for protein content, $r_{GS} = 0.38$ for grain yield and $r_{GS} = 0.16$ for protein yield.

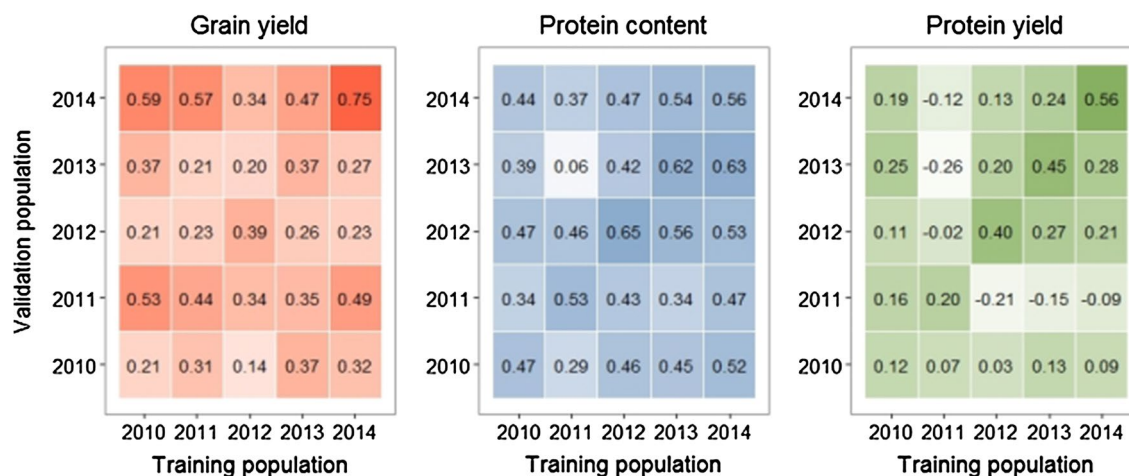
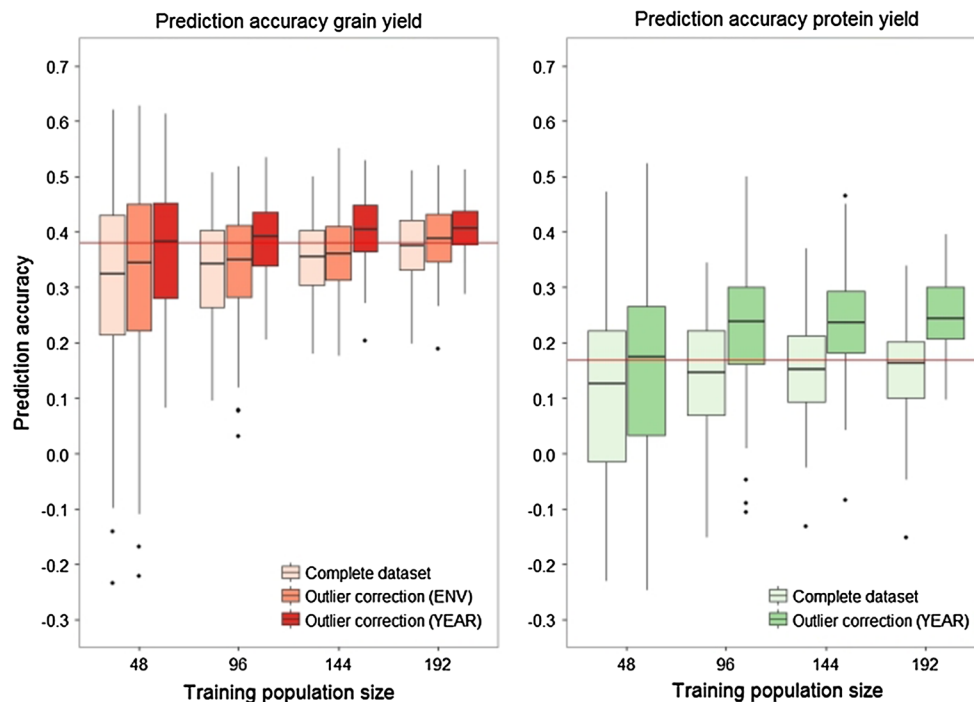


Fig. 2 Heatmap of the pair-wise prediction accuracy between breeding cycles on the off-diagonal and the result of the fivefold within-cycle cross-validation on the diagonal

Fig. 3 Influence of removing outlier years or environments from the training set on the prediction accuracy. Results were obtained using across-cycle cross-validation with years as folds. The horizontal red line indicates the maximum prediction accuracy in the complete dataset



Outlier correction and estimation of derived traits

Pair-wise prediction accuracies furthermore provided an approximation of the genetic correlation between breeding cycles. It was assumed that a breeding cycle with an overall low predictive ability also had a low value of representativeness for the breeding program across several breeding cycles, and was thus considered an outlier. The breeding cycle 2012 clearly presents itself as such an outlier with regard to grain yield. It achieved on average a much lower prediction accuracy when utilized as a training population

($r_{GS} = 0.26$) in comparison with all other breeding cycles ($r_{GS} = 0.36$) (Fig. 2). Although the heritability was relatively high for 2012, the phenotypic data was most likely strongly influenced by frost damage and drought stress in some trials. The low predictability and prediction accuracy furthermore identified the breeding cycle 2011 as an outlier for protein yield.

The influence of these outliers was investigated by omitting the above described breeding cycles when selecting lines for the training population. Using this approach, the prediction accuracy of protein yield increased from

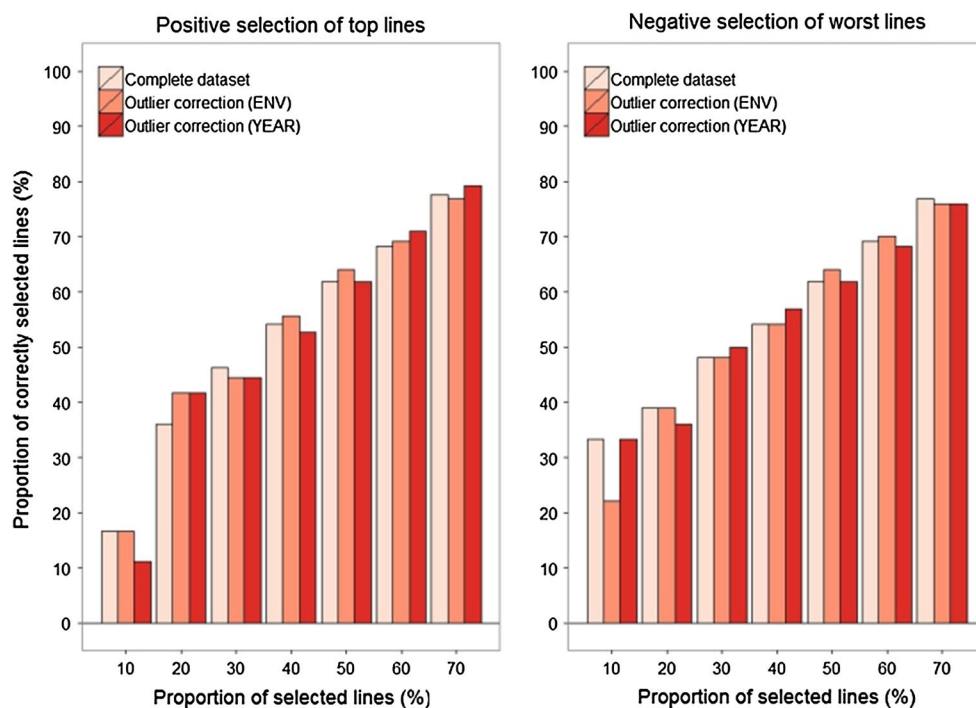


Fig. 4 Proportion of correctly selected lines when applying genomic selection for grain yield of either the best or worst lines in the independent validation population of the year 2015

$r_{GS} = 0.15$ to $r_{GS} = 0.25$ at a training population size of 192 lines (Fig. 3). A similar pattern was observed for grain yield at the same training population size, where the prediction accuracy raised from $r_{GS} = 0.38$ to $r_{GS} = 0.41$ when omitting phenotypic data from 2012. Strikingly a prediction accuracy of $r_{GS} = 0.41$ was estimated in the outlier corrected dataset using a training population size of 144 lines, surpassing the maximum of $r_{GS} = 0.38$ with a much larger training population of 256 lines in the complete dataset. One of the trials suffered severe drought stress and showed a negative phenotypic correlation with all other trials from 2012, and its mean prediction accuracy as a training population for all other trials was negative ($r_{GS} = -0.15$) and far below the average of all trials ($r_{GS} = 0.16$). Removing this trial from the phenotypic analysis of grain yield increased the heritability to $h^2 = 0.61$ and the average prediction accuracy of 2012 from $r_{GS} = 0.26$ to $r_{GS} = 0.30$, although the prediction accuracy obtained by cross-validation was not significantly higher than without outlier correction.

Grain yield was predicted with an accuracy of $r_{GS} = 0.34$ in the independent validation population of the breeding cycle 2015 ($h^2 = 0.57$). Dropping the drought stressed trial from the phenotypic analysis had no effect, while removing the entire breeding cycle 2012 decreased the prediction accuracy by 4 %. Removing 2011 from the training population decreased the prediction accuracy

of protein yield by 3 % in the independent validation ($h^2 = 0.30$). These minor changes in prediction accuracy had only a slight influence when genomically selecting for the top or against the worst lines tested in multi-environment trials in 2015 (Fig. 4).

The prediction accuracy of protein yield was rather low, while its component traits grain yield and protein content were estimated more accurately. The low estimated prediction accuracy for protein yield was raised from $r_{GS} = 0.16$ to $r_{GS} = 0.19$ by multiplying GEBVs of its component traits, instead of modeling protein yield directly, which corresponds to an increase of 19 %.

Discussion

Genomic selection has received attention in plant breeding research and caused some excitement in the last years (e.g. Heffner et al. 2009; Heslot et al. 2015). Nevertheless, results from practical applications in the framework of applied breeding programs are still sparse. This study focused on the problems and prospects of genomic selection in bread wheat. Five breeding cycles from an ongoing commercial breeding program were used as a base for assessing and enhancing the potential of genomic selection in bread wheat breeding.

Model selection

Since the introduction of genomic selection models from both the Bayesian (e.g. Technow and Melchinger 2013) and Frequentist methodology (Piepho 2009; Schulz-Streeck and Piepho 2010; Hofheinz and Frisch 2014) as well as machine learning methods (Ogutu et al. 2011, 2012) have been applied in plant breeding. Although great effort was put into developing these models no method showed clear superiority over the others across species or traits (Heslot et al. 2012). Hence we chose RR-BLUP as a computationally fast and robust alternative in our study.

Bias of the prediction accuracy

Genomic selection of non-phenotyped genotypes based on RR-BLUP is strongly dependent on the relationship between training population and selection candidates (Habier et al. 2007; Hayes et al. 2009). Empirical studies from plant breeding show a strong decline in accuracy when predicting distantly related populations (Riedelsheimer et al. 2013; Albrecht et al. 2014) and higher accuracies within closely related families (Lehermeier et al. 2014). Hence, the genetic relationship between training and selection population might introduce a bias in the estimation of prediction accuracy depending on the cross-validation scheme (Ly et al. 2013; Würschum et al. 2013). Genotype by environment interaction effects furthermore led to overestimations of the prediction accuracy, when genotypes from the training and selection population were tested in the same environment or year (Storlie and Charmet 2013; Krchov et al. 2015).

Both effects play important roles when predicting selection candidates across cycles in an ongoing breeding program. Accordingly, Hofheinz et al. (2012) reported an upward bias of the within-cycle prediction accuracy computed by cross-validation in comparison with the actual prediction accuracy across two subsequent breeding cycles. This observation was verified by our analysis and especially pronounced for protein yield, while the bias was much smaller for protein content. The highly heritable traits fusarium head blight resistance and plant height showed likewise less variation in the prediction accuracy across several breeding cycles, than the complex and low heritable trait grain yield in a dynamic barley breeding population (Sallam et al. 2015). A similar relationship between heritability and across-cycle prediction accuracy was also found in sugar beet (Hofheinz et al. 2012). Highly heritable traits are expected to have a less complex genetic architecture (Combs and Bernardo 2013), thus haplotype structures and relatedness responsible for the prediction accuracy (Daetwyler et al. 2010) might be preserved across breeding cycles. These considerations are in accordance with

the presented empirical results and suggest that prediction accuracy estimates for highly heritable traits are quite stable even across multiple breeding cycles.

Cross-validation results

Genomic selection is though especially interesting for low heritable traits and providing realistic estimates of trait-specific parameters is crucial for optimizing the resource allocations in an applied breeding program (Riedelsheimer and Melchinger 2013; Longin et al. 2015). Implementation of genomic selection in a breeding program faces the challenge of predicting a new set of genotypes with varying relatedness to previous generations or breeding cycles every year. Using breeding cycles as folds in cross-validation takes this problem into account, by sampling non-overlapping sets of genotypes from multiple breeding cycles as training populations and subsequent prediction of an independent breeding cycle. Estimates for grain yield derived from this across-cycle cross-validation scheme were on the upper bound of what has been reported before in mixed wheat populations (Heffner et al. 2011; Poland et al. 2012; Combs and Bernardo 2013; Storlie and Charmet 2013; Charmet et al. 2014; Isidro et al. 2015). Considering these studies together with our results a prediction accuracy between 0.3 (Longin et al. 2015) and 0.4 seems to be realistic for wheat grain yield across several breeding cycles. Interestingly these values correspond to the broad-sense heritability using variance components from Piepho et al. (2014) who analyzed long-term trends of bread wheat in the German official variety trials. A similar relationship between heritability and prediction accuracy across breeding cycles was previously observed by Hofheinz et al. (2012) for sugar content and molasses loss in sugar beet and several traits in barley (Sallam et al. 2015).

Particular with regard to the importance of phenotypic data (Bernal-Vasquez et al. 2014) a high estimate of the prediction accuracy was not unexpected in this study. First we selected trials with a high heritability from a larger population of target environments, as is common practice in plant breeding. Additionally, we selected only phenotypic records of lines that were tested orthogonally across all trials within a breeding cycle for building training populations. This allowed direct comparisons between all lines and consequently an expected higher efficiency than partial comparisons between lines or by using check varieties (Piepho et al. 2006). Even if a balanced subset cannot be extracted from existing data we recommend the use of the very best of trials to compute GEBVs for pending selection decisions as the data quality, measured by the heritability, is an important driving force of prediction accuracy in genomic selection (Jannink et al. 2010; Combs and Bernardo 2013).

Outlier correction and estimation of derived traits

A high heritability suggests that the phenotypic accuracy is high and gives good estimates of the underlying genotypes and true breeding values of the selection candidates (Visscher et al. 2008). Nevertheless, some years or trials have a low predictability despite of high heritability estimates (Dawson et al. 2013). Factors like biotic or abiotic stress caused by heavy rain, frost damage or drought can result in poor trial establishment and characterizes such environments as outliers (Heslot et al. 2013b). Although they give breeders the opportunity to select for special traits e.g. winter hardiness or resistance to a specific disease, dropping such outlier environments is justified when breeding for productivity and broad adaptation.

This outlier correction increased the prediction accuracy of grain yield, estimated by cross-validation, by 16 % in our study. We used pair-wise prediction accuracies and breeder's knowledge as an ad hoc measure to identify outliers for grain yield in wheat, though implementation of a systematic search algorithm led to analogous results for grain yield in barley (Heslot et al. 2013b). Dropping low-quality data for protein yield had a similar effect by raising the prediction accuracy by 50 % underpinning again the importance of phenotypic data. Independent validation suggests on the other hand that careful consideration is necessary before an outlier correction is undertaken. Dropping all phenotypic records of a genotype might even have a detrimental effect on the prediction accuracy in some cases as a broad genetic base and maximizing the phenotypic variance are essentials for optimizing a training population (Rincent et al. 2012; Isidro et al. 2015).

Apart from outlier correction another convenient option to improve the prediction accuracy for the derived trait protein yield was its prediction by component traits. The low prediction accuracy of protein yield could be slightly raised by multiplying GEBVs of the medium predictable traits grain yield and protein content. This approach might also be beneficial for other derived traits in plant breeding with a low heritability or prediction accuracy.

Conclusions

Numerous genomic selection studies were conducted in recent years, pointing out its large potential and several applied plant breeding programs adopted this new technology with high expectations. Hence results from multiple genomically selected breeding cycles are becoming available now, bringing these expectations to a realistic level. Genomic selection certainly opened up new opportunities by predicting difficult or expensive to phenotype traits or the estimation of derived traits by GEBVs of its components.

Furthermore the genomic selection framework helped to shed light on old problems, such as handling phenotypic data by approximating the genetic correlations among environments by their pair-wise prediction accuracy. Finally it also demands solutions to new problems such as optimizing training populations or redesigning breeding programs. Supported by the vast ongoing research, genomic selection is definitively becoming an integral part of modern bread wheat breeding and the future genetic improvement of crop plants.

Author contribution statement SM and CA analyzed the data and wrote the manuscript. HGR supported in the statistical analysis. FL, DE and HGU designed the field trials and collected the phenotypic data. FL and HB initiated and guided through the study. All authors read and approved the final manuscript.

Acknowledgments Open access funding provided by University of Natural Resources and Life Sciences Vienna (BOKU). We like to thank Maria Bürtstmayr and her team for the tremendous work when extracting the DNA of several hundred wheat lines each year as well as Herbert Hetzendorfer for managing the collection of the phenotypic data. This research was funded by the EU Eurostars projects “E! 6399 Genomic selection of wheat varieties for robustness, yield and quality” and “E! 8959 Genomic selection for nitrogen use efficiency in wheat”. We thank two anonymous reviewers for their valuable comments and suggestions for improving the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

Ethical standard The authors declare that the experiments comply with the current laws of Austria.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Albrecht T, Auinger HJ, Wimmer V, Ogutu JO, Knaak C, Ouzunova M, Piepho HP, Schön CC (2014) Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. *Theor Appl Genet* 127:1375–1386. doi:10.1007/s00122-014-2305-z
- Bernal-Vasquez AM, Möhring J, Schmidt M, Schönleben M, Schön CC, Piepho HP (2014) The importance of phenotypic data analysis for genomic prediction—a case study comparing different spatial models in rye. *BMC Genom* 15:646. doi:10.1186/1471-2164-15-646
- Bernardo R, Yu J (2007) Prospects for genomewide selection for quantitative traits in maize. *Crop Sci* 47:1082–1090. doi:10.2135/cropsci2006.11.0690

- Burgueno J, Cadena A, Crossa J, Banziger M, Gilmour AR, Cullis B (2000) User's guide for spatial analysis of field variety trials using ASREML. CIMMYT, Mexico, DF (Mexico). <http://repository.cimmyt.org/xmlui/handle/10883/584>. Accessed 16 Nov 2015
- Charmet G, Storile E, Oury FC, Laurent V, Beghin D, Chevarin L, Laupierre A, Perretant MR, Rolland B, Heumez E, Duchalais L, Goudmand E, Bordes J, Robert O (2014) Genome-wide prediction of three important traits in bread wheat. *Mol Breeding* 34:1843–1852. doi:[10.1007/s11032-014-0143-y](https://doi.org/10.1007/s11032-014-0143-y)
- Clark SA, Hickey JM, Datewyler HD, van der Werf JHJ (2012) The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genet Select Evol* 44:4. doi:[10.1186/1297-9686-44-4](https://doi.org/10.1186/1297-9686-44-4)
- Combs E, Bernardo RR (2013) Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. *Plant Genome* 6:1–7. doi:[10.3835/plantgenome2012.11.0030](https://doi.org/10.3835/plantgenome2012.11.0030)
- Crossa J, Pérez J, Hickey J, Burgueno J, Ornella L, Cerón-Roraz J, Zhang X, Dreisigacker S, Babu R, Li Y, Bonnet D, Mathews K (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* 112:48–60. doi:[10.1038/hdy.2013.16](https://doi.org/10.1038/hdy.2013.16)
- Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185:1021–1031. doi:[10.1534/genetics.110.116855](https://doi.org/10.1534/genetics.110.116855)
- Dawson JC, Endelman JB, Heslot N, Crossa J, Poland J, Dreisigacker S, Manès Y, Sorrells ME, Jannink JL (2013) The use of unbalanced historical data for genomic selection in an international wheat breeding program. *Field Crops Res* 154:12–22. doi:[10.1016/j.fcr.2013.07.020](https://doi.org/10.1016/j.fcr.2013.07.020)
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:1–10. doi:[10.1371/journal.pone.0019379](https://doi.org/10.1371/journal.pone.0019379)
- Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4:250–255. doi:[10.3835/plantgenome2011.08.0024](https://doi.org/10.3835/plantgenome2011.08.0024)
- Endelman JB, Jannink JL (2012) Shrinkage estimation of the realized relationship matrix. *G3* 2:1405–1413. doi:[10.1534/g3.112.004259](https://doi.org/10.1534/g3.112.004259)
- Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389–2397. doi:[10.1534/genetics.107.081190](https://doi.org/10.1534/genetics.107.081190)
- Habier D, Fernando RL, Garrick DJ (2013) Genomic BLUP decoded: a look into the black box of genomic prediction. *Genetics* 194:597–607. doi:[10.1534/genetics.113.152207](https://doi.org/10.1534/genetics.113.152207)
- Hayes BJ, Visscher PM, Goddard ME (2009) Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics research* 91:47–60. doi:[10.1017/S0016672308009981](https://doi.org/10.1017/S0016672308009981)
- Heffner EL, Jannink JL, Sorrells ME (2011) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* 4:65–75. doi:[10.3835/plantgenome2010.12.0029](https://doi.org/10.3835/plantgenome2010.12.0029)
- Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. *Crop Sci* 49:1–12. doi:[10.2135/cropsci2008.08.0512](https://doi.org/10.2135/cropsci2008.08.0512)
- Heslot N, Jannink JL, Sorrells ME (2013b) Using genomic prediction to characterize environments and optimize prediction accuracy in applied breeding data. *Crop Sci* 53:921–933. doi:[10.2135/cropsci2012.07.0420](https://doi.org/10.2135/cropsci2012.07.0420)
- Heslot N, Jannink JL, Sorrells ME (2015) Perspectives for genomic selection applications and research in plants. *Crop Sci* 55:1–12. doi:[10.2135/cropsci2014.03.0249](https://doi.org/10.2135/cropsci2014.03.0249)
- Heslot N, Rutkoski J, Poland J, Jannink JL, Sorrells ME (2013a) Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. *PLoS One* 8(9):e74612. doi:[10.1371/journal.pone.0074612](https://doi.org/10.1371/journal.pone.0074612)
- Heslot N, Yang HP, Sorrells ME, Jannink JL (2012) Genomic selection in plant breeding: a comparison of models. *Crop Sci* 52:146–160. doi:[10.2135/cropsci2011.09.0297](https://doi.org/10.2135/cropsci2011.09.0297)
- Hofheinz N, Borchardt D, Weissleder K, Frisch M (2012) Genome-based prediction of test cross performance in two subsequent breeding cycles. *Theor Appl Genet* 125:1639–1645. doi:[10.1007/s00122-012-1940-5](https://doi.org/10.1007/s00122-012-1940-5)
- Hofheinz N, Frisch M (2014) Heteroscedastic ridge regression approaches for genome-wide prediction with a focus on computational efficiency and accurate effect estimation. *G3* 4:539–546. doi:[10.1534/g3.113.010025](https://doi.org/10.1534/g3.113.010025)
- Huang YF, Poland JA, Wight CP, Jackson EW, Tinker NA (2014) Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat. *PLoS One* 9(7):e102448. doi:[10.1371/journal.pone.0102448](https://doi.org/10.1371/journal.pone.0102448)
- Isidro J, Jannink JL, Akdemir D, Poland J, Heslot N, Sorrells ME (2015) Training set optimization under population structure in genomic selection. *Theor Appl Genet* 128:145–158. doi:[10.1007/s00122-014-2418-4](https://doi.org/10.1007/s00122-014-2418-4)
- Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Briefings in functional genomics*. *Brief Funct Genomics* 9:166–177. doi:[10.1093/bfpg/eq001](https://doi.org/10.1093/bfpg/eq001)
- Krchov L-M, Gordillo GA, Bernardo R (2015) Multi-environment validation of the effectiveness of phenotypic and genome-wide selection within biparental maize populations. *Crop Sci* 55:1068–1075. doi:[10.2135/cropsci2014.09.0608](https://doi.org/10.2135/cropsci2014.09.0608)
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124(3):743–757
- Lehermeier C, Krämer N, Bauer E, Bauland C, Camisan C, Campo L, Flament P, Melchinger AE, Menz M, Meyer N, Moreau L, Moreno-González J, Ouzunova M, Pausch H, Ranc N, Schipprack W, Schönleben M, Walter H, Charcosset A, Schön CC (2014) Usefulness of multi-parental populations of maize (*Zea mays* L.) for genome-based prediction. *Genetics* 198:3–16. doi:[10.1534/genetics.114.161943](https://doi.org/10.1534/genetics.114.161943)
- Longin CF, Mi X, Würschum T (2015) Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. *Theor Appl Genet* 128:1297–1306. doi:[10.1007/s00122-015-2505-1](https://doi.org/10.1007/s00122-015-2505-1)
- Ly D, Hamblin MT, Rabbi I, Melaku G, Bakare M et al (2013) Relatedness and genotype × environment interaction affect prediction accuracies in genomic selection: a study in cassava. *Crop Sci* 53:1312–1325. doi:[10.2135/cropsci2012.11.0653](https://doi.org/10.2135/cropsci2012.11.0653)
- Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829. doi:[10.1534/genetics.117.007033](https://doi.org/10.1534/genetics.117.007033)
- Muir WM (2007) Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J Anim Breed Genet* 124:342–355. doi:[10.1111/j.1439-0388.2007.00700.x](https://doi.org/10.1111/j.1439-0388.2007.00700.x)
- Möhring J, Piepho HP (2009) Comparison of weighting in two-stage analysis of plant breeding trials. *Crop Sci* 49:1977–1988. doi:[10.2135/cropsci2009.02.0083](https://doi.org/10.2135/cropsci2009.02.0083)
- Ogut JO, Piepho HP, Schulz-Streeck T (2011) A comparison of random forests, boosting and support vector machines for genomic selection. *BMC Proc* 5(Suppl 3):S11. doi:[10.1186/1753-6561-5-S3-S11](https://doi.org/10.1186/1753-6561-5-S3-S11)
- Ogut JO, Schulz-Streeck T, Piepho HP (2012) Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. *BMC Proceedings* 6(Suppl 2):S10. doi:[10.1186/1753-6561-6-S2-S10](https://doi.org/10.1186/1753-6561-6-S2-S10)
- Piepho HP (2009) Ridge regression and extensions for genomewide selection in maize. *Crop Sci* 49:1165–1176. doi:[10.2135/cropsci2008.10.0595](https://doi.org/10.2135/cropsci2008.10.0595)

- Piepho HP, Büchse A, Truberg B (2006) On the use of multiple lattice designs and α -designs in plant breeding trials. *Plant Breeding* 125:523–528. doi:[10.1111/j.1439-0523.2006.01267.x](https://doi.org/10.1111/j.1439-0523.2006.01267.x)
- Piepho HP, Laidig F, Brobek T, Meyer W (2014) Dissecting genetic and non-genetic sources of long-term yield trend in German official variety trials. *Theor Appl Genet* 127:2599–2617. doi:[10.1007/s00122-014-2275-1](https://doi.org/10.1007/s00122-014-2275-1)
- Piepho HP, Möhring J (2007) Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* 177:1881–1888. doi:[10.1534/genetics.107.074229](https://doi.org/10.1534/genetics.107.074229)
- Piyasatian N, Fernando RL, Dekkers JCM (2007) Genomic selection for marker-assisted improvement in line crosses. *Theor Appl Genet* 115:665–674. doi:[10.1007/s00122-007-0597-y](https://doi.org/10.1007/s00122-007-0597-y)
- Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y et al (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5:103–113. doi:[10.3835/Plantgenome2012.06.0006](https://doi.org/10.3835/Plantgenome2012.06.0006)
- R development core team (2015) R: a language and environment for statistical computing. <http://www.r-project.org/>
- Riedelsheimer C, Endelman JB, Stange M, Sorrells ME, Jannink JL, Melchinger AE (2013) Genomic predictability of interconnected biparental maize populations. *Genetics* 194:493–503. doi:[10.1534/genetics.113.150227](https://doi.org/10.1534/genetics.113.150227)
- Riedelsheimer C, Melchinger AE (2013) Optimizing the allocation of resources for genomic selection in one breeding cycle. *Theor Appl Genet* 126:2835–2848. doi:[10.1007/s00122-013-2175-9](https://doi.org/10.1007/s00122-013-2175-9)
- Rincent R, Laloë D, Nicolas S, Altmann T, Brunel D et al (2012) Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.). *Genetics* 192:715–728. doi:[10.1534/genetics.112.141473](https://doi.org/10.1534/genetics.112.141473)
- Rutkoski J, Singh RP, Huerta-Espino J, Bhavani S, Poland J, Jannink JL, Sorrells ME (2015) efficient use of historical data for genomic selection: a case study of stem rust resistance in wheat. *Plant Genome* 8:1. doi:[10.3835/plantgenome2014.09.0046](https://doi.org/10.3835/plantgenome2014.09.0046)
- Saghai-Marouf MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphism in barley: mendelian inheritance, chromosomal location, and population dynamics. *PNAS* 81:8014–8019. doi:[10.1073/pnas.81.24.8014](https://doi.org/10.1073/pnas.81.24.8014)
- Sallam A, Endelman J, Jannink JL, Smith K (2015) Assessing genomic selection prediction accuracy in a dynamic barley breeding population. *Plant Genome* 8:1. doi:[10.3835/plantgenome2014.05.0020](https://doi.org/10.3835/plantgenome2014.05.0020)
- Schulz-Streeck T, Oguto JO, Piepho HP (2011) Pre-selection of markers for genomic selection. *BMC Proc* 5(Suppl 3):S12. doi:[10.1186/1753-6561-5-S3-S12](https://doi.org/10.1186/1753-6561-5-S3-S12)
- Schulz-Streeck T, Piepho HP (2010) Genome-wide selection by mixed model ridge regression and extensions based on geostatistical models. *BMC Proc* 4(Suppl 1):S8. doi:[10.1186/1753-6561-4-S1-S8](https://doi.org/10.1186/1753-6561-4-S1-S8)
- Storlie E, Charmet G (2013) Genomic selection accuracy using historical data generated in a wheat breeding program. *Plant Genome* 6:1–9. doi:[10.3835/plantgenome2013.01.0001](https://doi.org/10.3835/plantgenome2013.01.0001)
- Technow F, Melchinger AE (2013) Genomic prediction of dichotomous traits with Bayesian logistic models. *Theor Appl Genet* 126:1133–1143. doi:[10.1007/s00122-013-2041-9](https://doi.org/10.1007/s00122-013-2041-9)
- VSN International (2015) ASReml 3. VSN International, Hemel Hempstead
- VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91:4414–4423. doi:[10.3168/jds.2007-0980](https://doi.org/10.3168/jds.2007-0980)
- Visscher PM, Hill WG, Wray NR (2008) Heritability in the genomics era—concepts and misconceptions. *Nat Rev Genet* 9:255–266. doi:[10.1038/nrg2322](https://doi.org/10.1038/nrg2322)
- Würschum T, Abel S, Zhao Y (2014) Potential of genomic selection in rapeseed (*Brassica napus* L.) breeding. *Plant Breeding* 133:45–51. doi:[10.1111/pbr.12137](https://doi.org/10.1111/pbr.12137)
- Würschum T, Reif JC, Kraft T, Janssen G, Zaho Y (2013) Genomic selection in sugar beet breeding populations. *BMC Genet* 14:85. doi:[10.1186/1471-2156-14-85](https://doi.org/10.1186/1471-2156-14-85)
- Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP, Longin FH, Rank N, Reif JC (2012) Accuracy of genomic selection in European maize elite breeding populations. *Theor Appl Genet* 124:769–776. doi:[10.1007/s00122-011-1745-y](https://doi.org/10.1007/s00122-011-1745-y)